Charles University
Faculty of Science

Study program: Special Chemical and Biological Programmes

Branch of study: Molecular Biology and Biochemistry of Organisms

Brenda Tomčalová

Effects of phosphorylation on 3D conformation of proteins

Vliv fosforylace na 3D konformaci proteinů

Bachelor's thesis

Supervisor: Mgr. Marian Novotný Ph.D.

Prague, 2017

Acknowledgements

I would like to express my sincere gratitude to my supervisor Marian Novotný. Without his suggestions, support and understanding this bachelor thesis would have never been accomplished.

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze dne

Podpis: ……

Brenda Tomčalová

Abstrakt

Fosforylace je jedna z nejrozšířenějších proteinových posttranslačních modifikací. Pochopení funkce a regulace fosforylace je významné pro diagnózu a léčbu mnoha onemocnění. Tato práce shrnuje studie analyzující fosforylace pomocí bioinformatických nástrojů na velkých datových souborech. V této práci jsou popsány obecné principy, jak fosforylace ovlivňuje fyzikálně-chemické vlastnosti proteinů. V prvních dvou kapitolách jsou nastíněny evoluční zákonitosti, funkce a regulace proteinových kináz a fosfatáz. Ve třetí kapitole se práce zaměřuje na rozložení fosforylačních míst proteinů napříč různými organizmy (rostlinnými a živočišnými). Poslední kapitola se věnuje shrnutí dnešních poznatků o alosterickém a ortosterickém efektu fosforylace a vlivu fosforylace na 2D a 3D strukturu proteinů.

Klíčová slova: fosforylace, terciální struktura, sekundární struktura, neuspořádané oblasti

Abstract

Phosphorylation is one of the most ubiquitous posttranslational modification types. Understanding of its function and regulation has significant impact on diagnosis and treatment of many diseases. This thesis presents and summarizes the results of several publications that analyze phosphorylation on large datasets using bioinformatics tools. In this thesis the general principles how the phosphorylation influences physico-chemical properties of proteins are described. In the first and the second chapter the evolution principles, function and regulation of protein kinases and phosphatases are provided. In the third chapter thesis concentrates on the distributions of phosphorylated sites among organisms (plants and animals). In the last chapter current knowledge of orthosteric and allosteric effects of phosphorylation as well as its effects on 2D and 3D structure of phosphorylated proteins is summarized.

Key words: phosphorylation, tertiary structure, secondary structure, disordered regions

List of abbreviations

ATP = adenosine triphosphate

GTP = guanosine triphosphate

ERK2 = extracellular regulated kinase 2

pSer = phosphorylated serine

pThr = phosphorylated threonine

PTM = posttranslational modification

PTP = protein tyrosine phosphatase

pTyr = phosphorylated tyrosine

SH2 domain = Src-homology 2 domain

SH3 domain = Src-homology 3 domain

Content

1. Introduction

Phosphorylation represents one of the most ubiquitous posttranslational modification (PTM) types in eukaryotes, prokaryotes and archaea. It is a reversible posttranslational modification of proteins, in which a protein kinase covalently adds a phosphate group, consisting of one phosphorous and four oxygen, to one of substrate amino acid side chains. In addition, phosphorylation changes physico-chemical properties, stability, kinetics and dynamics of protein. Phosphorylation can, therefore, modulate the ability of a protein to perform chemical reactions (phosphorylation can result either in activation and inhibition of enzyme activity) and to interact with other proteins (Cohen, 2000).

Phosphorylation regulates many essential biological processes including gene transcription and translation regulation, membrane transport, hormonal response, muscle contraction, light harvesting and photosynthesis, cell division, growth or viral oncogene response (Johnson and Lewis, 2001). Therefore, a detailed understanding of phosphorylation events and their regulation is crucial to fully grasp the biology of many fundamental cellular processes. In addition, the understanding of function and regulation phosphorylated proteins has significant impact on diagnosis and treatment of many illnesses. For example, tyrosine kinase inhibitors such as Gleevec, Iressa and Tarceva are used to treat cancers (e.g. chronic myelogenous leukemia, acute lymphocytic leukemia, gastrointestinal stromal tumors or systemic mastocytosis) (Paul and Mukhopadhyay, 2004).

Although phosphorylation is of a great interest, many phosphorylated sequences and structures still remain undiscovered. Aim of this thesis is to summarize works that describe and analyze conformational changes in proteins upon phosphorylation in a high-throughput manner.

2. Phosphorylation

2.1. Basic principles of phosphorylation

Protein phosphorylation is an enzyme-catalyzed attack of nucleophilic hydroxyl or amino group on the gamma-phosphate group (γ-PO_3^{2-}) of an adenosine triphosphate (ATP) or guanosine triphosphate (GTP). This reaction is facilitated by Mg^{2+} and catalyzed by a protein kinase (Figure 1).

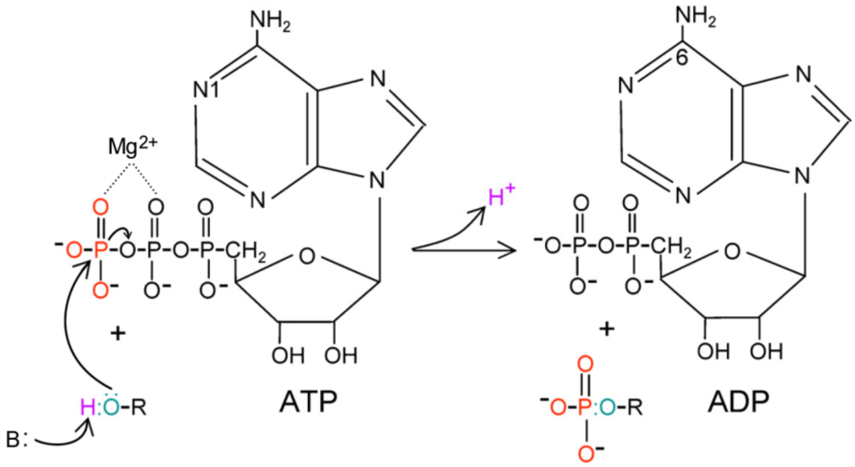

Figure 1: Mechanism of nucleophilic attack of base resulting in transfer of phosphate group from ATP on hydroxyl group of protein. Adapted from Roskoski, "ERK1/2 MAP kinases: Structure, function, and regulation"(Roskoski, 2012).

Protein kinases bind ATP/GTP and substrate proteins and transfer the gamma phosphate group from ATP/GTP to amino acids with free hydroxyl (-OH) groups such as serine, threonine and tyrosine. The products of the reaction are the phosphorylated protein and adenosine or guanine diphosphate (ADP/GDP) (Figure 2).

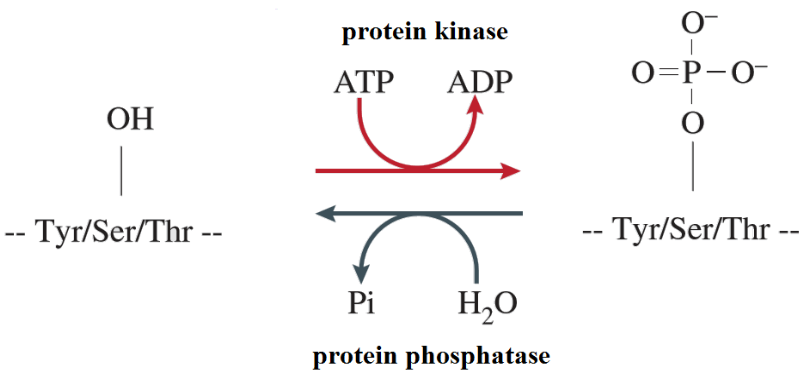

Figure 2: Phosphorylation reaction catalyzed by protein kinases. The amino acids with free hydroxyl (-OH) groups such as serine, threonine and tyrosine of certain proteins are modified with the gamma phosphate from ATP. The opposite reaction to phosphorylation is called dephosphorylation and is catalyzed by protein phosphatases. Adapted from Jin & Pawson, "Modular evolution of phosphorylation-based signalling systems" (Jin and Pawson, 2012).

In many cases, phosphorylation on proteins occurs more than once and this process is called multiple phosphorylation. Phosphorylation is essentially unidirectional because of the large amount of free energy released by the breakage of phosphodiester bond in ATP. Kinases can also catalyze the reverse reaction in the excess of ADP and regenerate OH^- group and remove phosphate from the phosphorylated residue in a target protein to generate ATP because of relatively high bond energy of phosphate ester linkage (Hunter, 2010)(Hunter, 2012). It is, however, uncommon in biological processes in the cell. More frequently the hydrolysis and breakage of phosphodiester bonds is catalyzed by a protein phosphatase. This opposite reaction to phosphorylation is called dephosphorylation.

Phosphorylation is often associated with other posttranslational modifications (PTMs) such as acetylation, ubiquitination or methylation (Figure 3) (Duan and Walther, 2015).

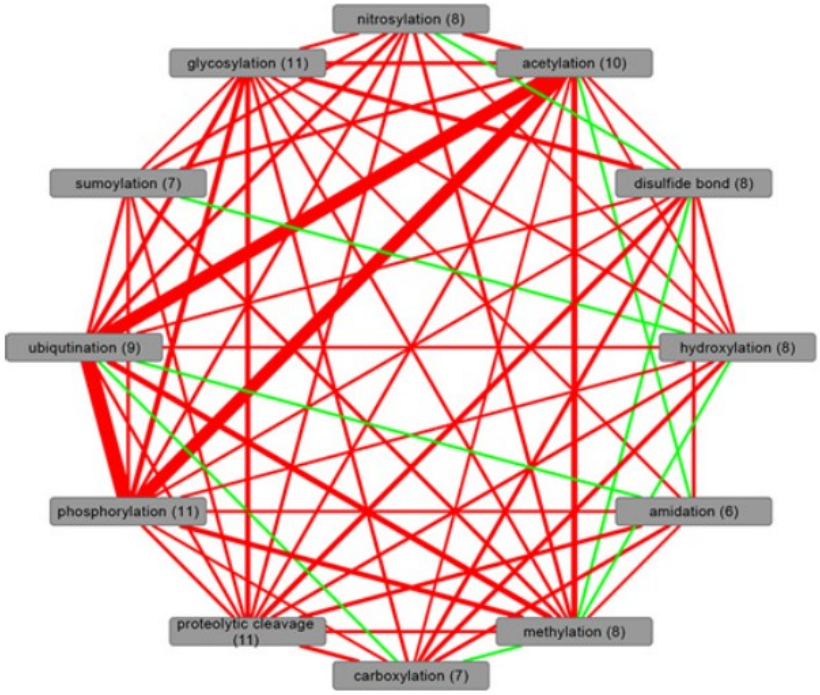

Figure 3: Co-existence network of posttranscriptional modifications in the *Homo sapiens* proteome. It illustrates how phosphorylation is connected with other types of posttranscriptional modifications such as methylation, ubiquitination or acetylation. Adopted from Duan & Walther, „The Roles of Post-translational Modifications in the Context of Protein Interaction Networks“ (Duan and Walther, 2015).

When one PTM serves as a signal for the addition or removal of a second PTM or for recognition by a binding protein that carries out a second modification, crosstalk is called “positive”. Reversely when one PTM masks the recognition site for second PTM or when one PTM competes with another one, crosstalk is called “negative”. In principle, cells can precisely

response to stimuli because different PTMs can positively or negatively regulate the interaction of domains, binding of substrate or recognition of another PTMs (Hunter, 2007).

2.2. Properties of phosphorylated amino acids

Phosphorylation of nine amino acids are chemically feasible: serine, threonine, tyrosine, arginine, lysine, histidine, cysteine, aspartate and glutamine (Figure 4). Study of phosphorylated serine, threonine and tyrosine is more common than phosphorylation of other amino acids because of their long life-time and acid stability of the phosphodiester bond. In addition, phosphorylation of these tree amino acids occurs in eukaryotes and thus the dedicated tools such as specific antibodies are more researched (Cieśla et al., 2011).

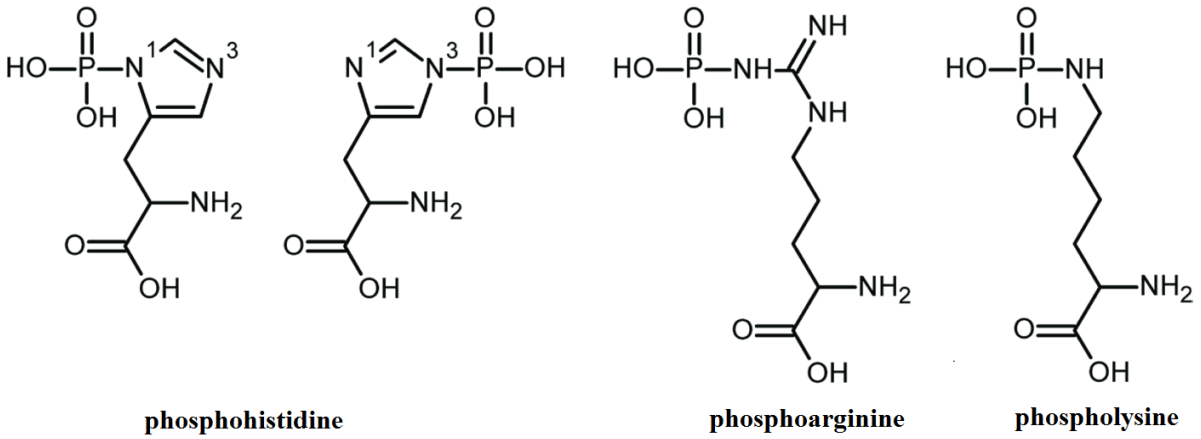

Figure 4: Examples of phosphorylated basic amino acids with phosphoramidate bond. Structures of N-1-phosphohistidine, N-3-phosphohistidine, N^{ω}-phosphoarginine and N^{ε}-phospholysine. Adapted from Cieśla et al., "Phosphorylation of basic amino acid residues in proteins" (Cieśla et al., 2011).

Out of these three phosphorylated residues, phosphorylation on tyrosine is the most difficult to study. Phosphorylated tyrosine is typically short-living owing to the presence of extremely active specific tyrosine phosphatases (PTPs) that rapidly dephosphorylate any phosphorylated tyrosine residue that is not protected through binding to a SH2, PTP domain or via an intramolecular interactions. As a result, phosphorylated tyrosine constitutes only 1 % of the total phosphorylated amino acids in proteins in a typical mammalian cell (Hunter, 2014).

It is important to mention that phosphorylated serine, threonine and tyrosine have each special properties. Unlike serine, threonine has the methyl group on the Cβ atom of the threonine side chain, which controls the side-chain flexibility and the relative population of

fully solvated phosphate group (Figure 5). Therefore phosphorylated serine and threonine have different intramolecular H-bonding propensity of the side chain phosphoryl group.

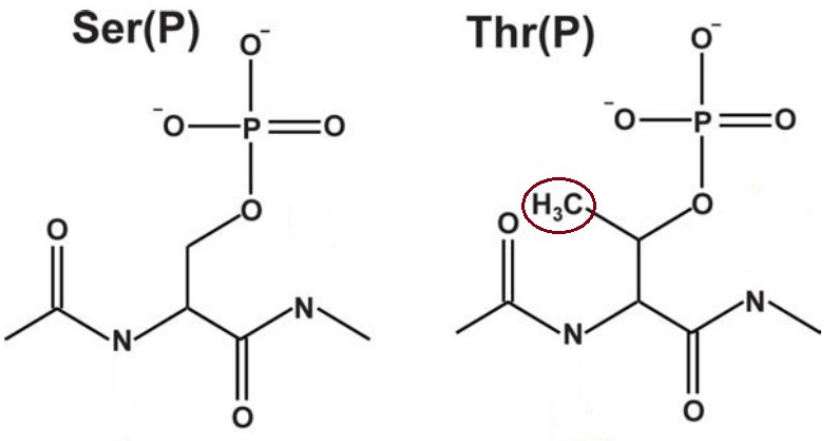

Figure 5: Molecular structures of phosphorylated serine and threonine peptides. Adapted from Kim et al., „Phosphorylation alters backbone conformational preferences of serine and threonine peptides“ (Kim et al., 2011).

Phosphorylated tyrosine is also specific in its properties. Phosphate on tyrosine is linked to the O^4 position of the phenolic ring, so it lies much further away from the peptide backbone than the phosphate on the β-OH groups of serine and threonine. Therefore, the “reader” domains which bind phosphorylated tyrosine (e.g. SH2 domain) have deeper binding pockets than those for binding phosphorylated serine and threonine. It ensures the specificity of binding the phosphorylated tyrosine residues. In addition, the phenolic ring of tyrosine could mediate more hydrophobic or π bond-ring interactions than could mediate serine or threonine (Hunter, 2014).

2.3. Databases of protein phosphorylation

Information about experimentally verified phosphorylations is collected and collated by phosphorylation databases, which often including also data of the other posttranslational modifications. The most used phosphorylation databases are presented in Table 1.

Table 1: Databases of protein phosphorylation. The first column contains the names of databases, the second column where could be available and third the reference.

Database	Website	Reference
PhosphoSitePlus (PSP)	http://www.phosphosite.org/	(Hornbeck et al., 2012)
Phospho.ELM	http://phospho.elm.eu.org/	(Dinkel et al., 2011)
Phosphorylation site database (PHOSIDA)	http://141.61.102.18/phosida/	(Gnad et al., 2011)
Human Protein Reference Database (HPRD)	http://www.hprd.org/	(Goel et al., 2012)
PhosphoPep	http://www.phosphopep.org/	(Bodenmiller et al., 2009)
MtcPTM	retired[1]	Not published
Universal protein resource (UniProt)	http://www.uniprot.org/	(Apweiler et al., 2014)
Human Proteinpedia	http://www.humanproteinpedia.org/	(Goel et al., 2012)
PhosphoBase	nowadays integrated by Phospho.ELM : http://phospho.elm.eu.org/	(Kreegipuu et al., 1999)
dbPTM	http://dbptm.mbc.nctu.edu.tw/	(Huang et al., 2016)

[1] MtcPTM database was an online repository of human and mouse phosphosites since 2007. It was sorted by biologically relevant experimental information such as used methods and experiment conditions. Database doesn't exist and isn't a part of any existing database.

3. Protein kinases and phosphatases

Protein kinases and phosphatases are fundamental for basic signaling machinery mediated by phosphorylation. Based upon the nature of the phosphorylated hydroxyl group of substrate, protein kinases are often classified as serine/threonine kinases, tyrosine kinases and tyrosine-kinase like proteins. The protein serine/threonine kinases catalyze the transfer of the terminal γ-phosphate group from ATP/GTP to serine/threonine hydroxyl groups on proteins and the protein tyrosine kinases to tyrosine hydroxyl groups on proteins.

Kinases have evolutionary conserved protein kinase domain (Figure 6)(Huse and Kuriyan, 2002). Serine/threonine kinase domain was first described by Knighton et al. for adenosine monophosphate-dependent protein kinase (Knighton et al., 1991).

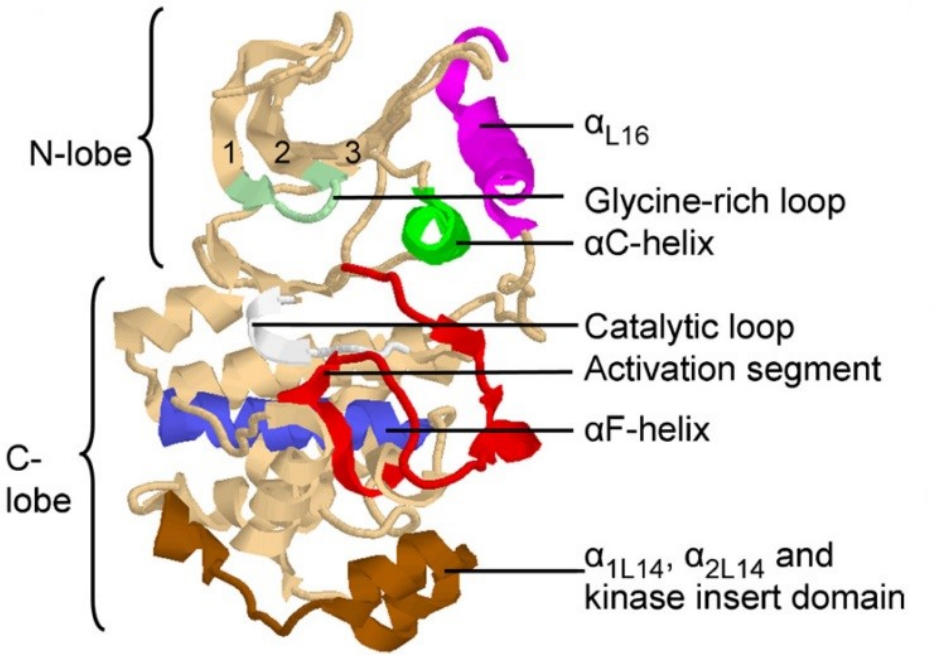

Figure 6: Ribbon diagram of human extracellular regulated kinase 2 (ERK2) in inactive state. ERK2 is a typical serine/threonine kinase with conserved protein kinase domain similar to adenosine monophosphate-dependent protein kinase. Adapted from Roskoski, „ERK1/2 MAP kinases: Structure, function, and regulation“ (Roskoski, 2012).

A characteristic bilobal structure includes a smaller N-terminal lobe, rich in antiparallel β sheets (β1-5) and containing a conserved αC-helix, and a bigger C-terminal lobe, that is mostly α helical. The catalytic site for ATP is located between these two lobes. ATP is bound to the small lobe of the kinase with β and γ phosphates positioned under the glycine-loop with motif Gly-X-Gly-X-X-Gly (sometimes called P-loop), where X is any amino acid. Glycine rich P-loop is positioned between β1 and β2 sheet of the small N-terminal lobe and is the most

flexible part of the N-lobe. Protein substrate is positioned in the active site cleft between small and big lobe and associated with the activation loop of the C-terminal lobe (Knighton et al., 1991).

3.1. Evolution of phosphorylation

Prokaryotes and simple eukaryotes such as bacteria, some fungi and plants use phosphorylation on histidine, cysteine and aspartate for regulations of signaling cascades. In contrast, phosphorylation in high eukaryotes preferentially occurs on serine, threonine and tyrosine. However, it is not strictly defined. For instance dual-specific kinases, which mediate phosphorylation on serine, threonine and tyrosine, were found to phosphorylate tyrosine in pre-metazoans. However, tyrosine kinases were thought to be a Metazoan invention. In 2008 Pincus et al. used the SMART domain resource to identify tyrosine kinases, phosphatases and SH2 domain containing proteins in 41 published eukaryotic genome sequences. Evolution of phosphorylation on tyrosine was found important for transition to metazoan multicellularity, because of the signal encoding potential that was available by evolution new, modular reader/writer/eraser signaling system (Pincus et al., 2008).

Tyrosine-phosphorylated proteins were found also in the genomes of choanoflagellates. Choanoflagellates, close relatives of metazoans, include the choanoflagellate *Monosiga brevicollis*, which has at least 43 tyrosine kinases, 34 tyrosine phosphatases. Further King et al. found that the genome of *M. brevicollis* contains also 100 proteins with SH2 domain. This domain was suggested to be invented in evolution for recognition and protection of unstable phosphorylated tyrosine residues. (King et al., 2008).

Pincus et al. in 2008 suggested that phosphorylation tyrosine signaling machinery emerged only shortly before the divergence of metazoans and choanoflagellates, shortly before the evolution of metazoan multicellularity. Tyrosine signaling has, therefore, evolved in choanoflagellates differently from metazoans (Figure 7) (Pincus et al., 2008).

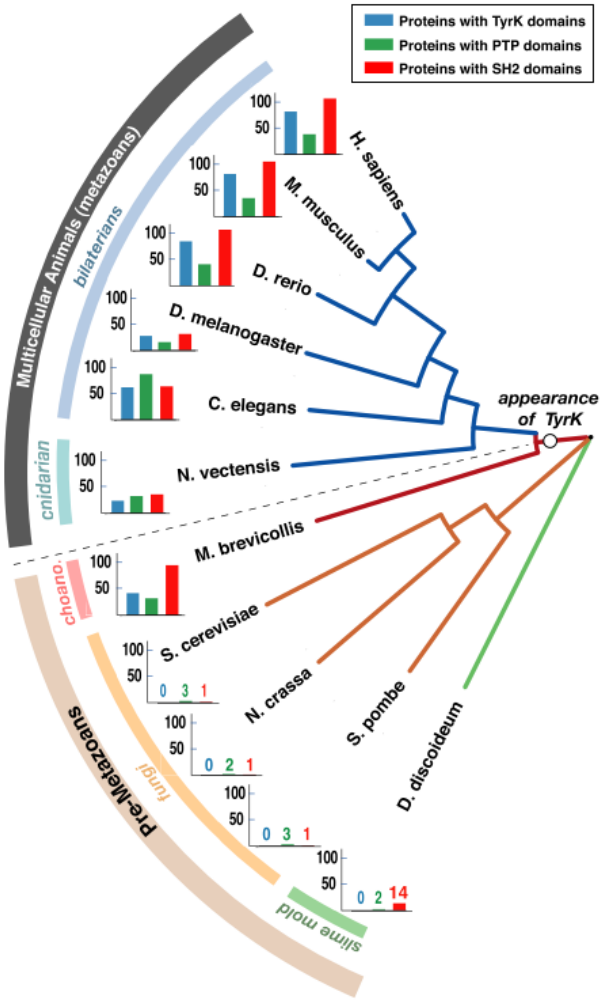

Figure 7: Scheme shows number of proteins containing tyrosine kinases (blue), phosphatases (green) and/or SH2 domains (red) among different eukaryotes. Only choanoflagellates and metazoans have high numbers of all three domains. Adopted from Pincus et al., „Evolution of the phospho-tyrosine signaling machinery in premetazoan lineages“ (Pincus et al., 2008).

Although the phosphorylation on tyrosine is the key mechanisms to cell-cell communication in metazoans, the content of phosphorylated tyrosine negatively correlates with the number of distinct cell types in metazoan species. Generally, the more complicated body structure an organisms has, the fewer phosphorylated tyrosine within its proteins are observed. On the contrary, the number of genomically encoded tyrosine kinases increases with biological complexity. According to Tan et al., it is probably a result of the optimization of newly emerged signaling networks through elimination of unspecific detrimental phosphorylation events by tyrosine-removing mutations (Tan et al., 2009). In 2015 Pandaya et al. tested the hypothesis that the metazoan tyrosine loss was driven by selection against promiscuous phosphorylation events and found that tyrosine loss in evolution is neutral with respect to phosphorylation. (Pandya et al., 2015)

3.2. Roles of protein kinases and phosphatases in signalization

Kinases are characteristic for a low affinity and a high specificity of substrate binding, which make kinases ideal for signaling. As a result, protein kinases and phosphatases play pivotal roles in regulating and coordinating such diverse processes as cell adhesion, proliferation, differentiation, migration, survival apoptosis (Wortzel and Seger, 2011), regulation of the eukaryotic cell division cycle and transcription, mRNA processing, metabolism, stem cell self-renewal, the differentiation of nerve cells and spermatogenesis (Enserink and Kolodner, 2010)(Lim and Kaldis, 2013).

Single protein kinases can phosphorylate one specific or multiple phosphorylation sites. One phosphorylation site can be simultaneously a target of many different protein kinases. The same is true for protein phosphatases. This mechanism enables regulation of signaling pathways. Moreover many signaling pathways have multiple crosstalk pathways points. Therefore, phosphorylation of a constituent proteins of one pathway ("junction proteins") may affect the function of another pathway (Figure 8) (Nishi et al., 2011).

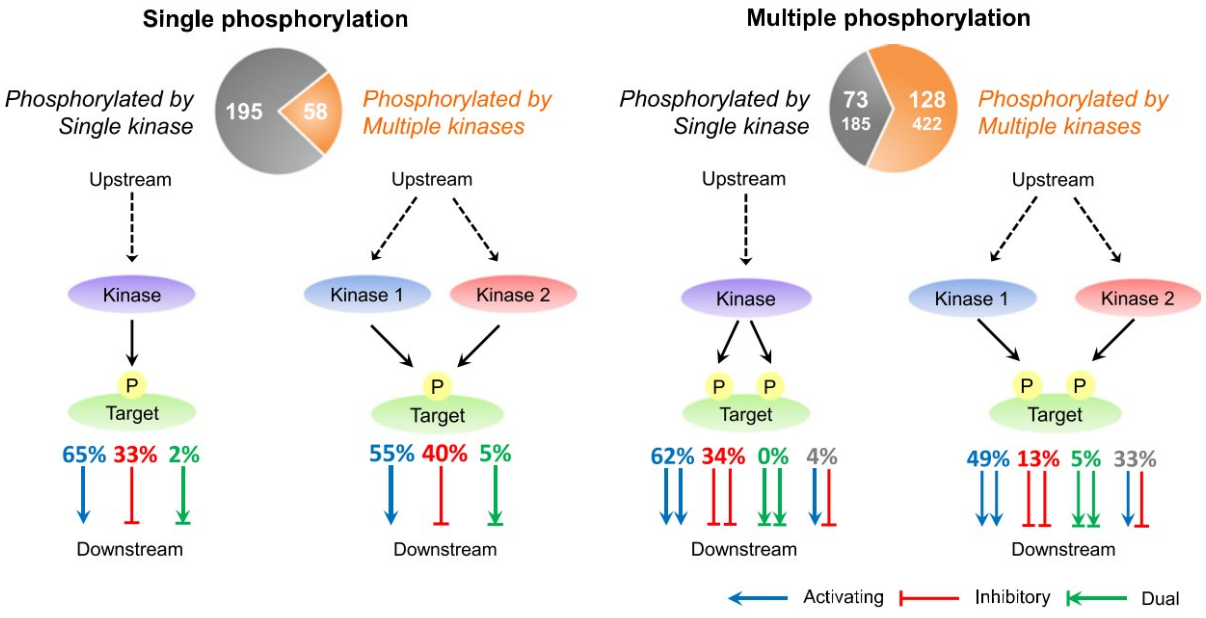

Figure 8: Illustration of pathway regulation by single or multiple phosphorylation. Above is shown if single or multiple phosphorylation is caused by single (gray) or multiple (orange) kinases and downstream is pointed the effect on a target protein and is illustrated in percentages how many phosphorylated targets ("junction proteins") may provide pathway crosstalk by another pathway activation, inhibition or both. Adopted from Nishi et al., „Crosstalk between signaling pathways provided by single and multiple protein phosphorylation sites" (Nishi et al., 2015).

Nishi et al. observed in 2015 that the proteins which are phosphorylated by only one kinase participated in a smaller number of pathways than proteins phosphorylated by multiple kinases. It is probably because of selective evolution. Proteins which are regulated by many different kinases could be preferred because they are better regulated (Nishi et al., 2015).

3.3. Phosphorylation motifs

Protein kinases recognize phosphorylation sites with specific residues in a neighborhood, which are together called phosphorylation motifs. Phosphorylation motifs can be specific for organisms and even for tissues.

In 2016 Karabulut & Frishman studied tissue-specific preferences of phosphorylation sites. They used the dataset of 3 1480 phosphorylation sites in 7 280 proteins identified by high-resolution tandem mass spectrometry in 14 rat tissues: brain (dissected into cerebellum, cortex and brainstem), heart, muscle, lung, kidney, liver, stomach, pancreas, spleen, thymus, intestine, testis, perirenal fat, and blood. They found that protein kinases as well as spatial motifs of phosphorylation are tissue-specific. They, further, observed that global phosphorylation sites of serine, threonine and tyrosine have commonly in their upstream regions negatively charged glutamic acid, aspartic acid or polar serine. They also showed that proline residues at position +1, charged lysine and arginine residues (except for the positions from +1 to +4) are typically in the neighborhood of phosphorylated serine and threonine (Karabulut and Frishman, 2016).

In 2015 Frades et al. analyzed phosphorylation patterns across eukaryotes and published that phosphorylation motifs and kinases are evolutionary specific for each kingdoms/phyla. For example whereas animals have typically arginine residues on both the N- and C- terminal sides of phosphoserines, plants have on the C-terminal side acidic residues (Frades et al., 2015).

4. Distribution of phosphorylated residues

In 2006, Olsen et al. used liquid chromatography-mass spectrometry and detected 2 000 phosphopeptides from HeLa cells stimulated by epidermal growth factor. Further they identified 103 phosphorylated tyrosine, 670 phosphorylated threonine, and 4 901 phosphorylated serine. They found the distribution of phosphorylated serine, threonine and tyrosine is 86,4 %, 11,8 % and 1,8 %. Further they measured the distribution of single, doubly, triply and quadruply phosphorylated peptides in their dataset and showed in Figure 9 that cirka 32,5 % of used phosphorylation sites are multiply phosphorylated. It suggested that many phosphorylated peptides could have been phosphorylated more times. Unfortunately, the numbers for each category aren't accurate because the categories are overlapping (Olsen et al., 2006).

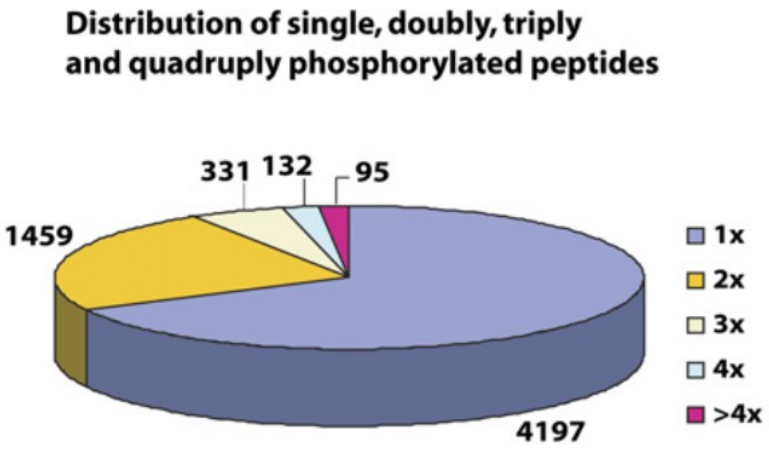

Figure 9: Distribution between singly and multiply phosphorylated peptides. Adopted from Olsen et al., „Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks" (Olsen et al., 2006).

Nishi et al. used data set including 382 homooligomers and 551 heterooligomers with 1 983 phosphorylation sites altogether adapted from phosphorylation databases PhosphoSitePlus, Phospho.ELM and PHOSIDA. They analyzed this non-redundant data set of phosphorylated proteins and noted distribution of phosphorylated residues 40 % for serine, 25 % for threonine and 35 % for tyrosine for human protein structural complexes. (Nishi et al., 2011).

Song et al. in 2012 collected experimentally identified phosphorylation sites from databases containing proteins with known phosphorylation sites, such as PhosphoPep (version 2.0), Phospho.ELM 8.3 (released in April 2010), SysPTM 1.1, PhosphoSitePlus, and Human Protein Reference Database 9.0. They collected also thousands of phosphorylation sites from several

articles containing experimentally gained data. Their dataset included 145 646 phosphorylation sites in 28 457 substrates.. There were 14 534 phosphorylation sites from *S. cerevisiae* , 5555 from *C. elegans* , 15 622 from *D. melanogaster*, 49 119 from *M. musculus*, and 60 816 phosphorylation sites from *H. sapiens*. Than the team of Song observed phosphorylation sites from the non-tumor human liver tissue. They used reversed-phase-reversed-phase liquid chromatography, high-throughput mass spectrometry and a new data process platform of ArMone to conduct a large-scale phosphorylation analysis of the human liver. They identified 10 644 nonredundant phosphopeptides, including 7 214 (67,8 %) singly, 2 403 (22,6 %) doubly and 1 027 (9,6 %) triply phosphorylated peptides. So their analysis contained over 30 % multiply phosphorylated peptides (Figure 10).

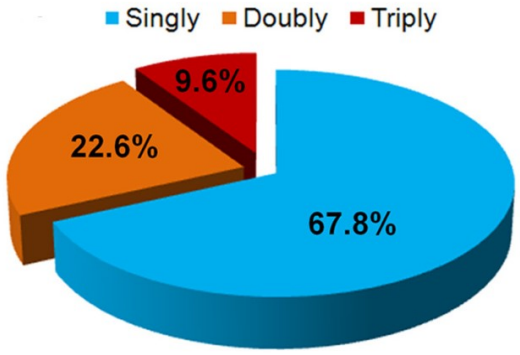

Figure 10: Distribution of singly, doubly and triply phosphorylated peptides in human liver. Adapted from Song et al., „Systematic analysis of protein phosphorylation networks from phosphoproteomic data“ (Song et al., 2012).

Further they measured the distribution of 9 719 phosphoryltion sites in the 2 998 phosphopeptides from UniProt database. Distribution of phosphoserine, phosphothreonine and phosphotyrosine is 85,3 %, 12,9 % and 1,8 %. Song et al. analyzed also whole human phosphoproteome and got the distribution of 59,8 % for phosphoserine, 17,6 % for phosphothreonine and 22,6 % for phosphotyrosine (Figure 11). Distribution of phospho-rylated residues in human liver is in this case different from the distribuiton in the whole human phosphoproteome. Song et al. thought, it could indicate that distribution of phosphorylation is also tissue-specific. However, they compared large datasets gained from different databases, which contained phosphoproteins acquired by different experimental methods in different points of the cell life, with their results gained by their experimental work. In fact, phosphorylated tyrosine is less stable in phosphoamino acid analysis than phosphorylated serine and threonine. Whereas the results from the human liver presented by

the team of Song could be biased, the phosphorylation databases collect data from many works and so contain more identified phosphorylated tyrosine. Therefore it is natural that the distribution of phosphorylated reisdues in databases differs from experimentally gained distribution (Song et al., 2012).

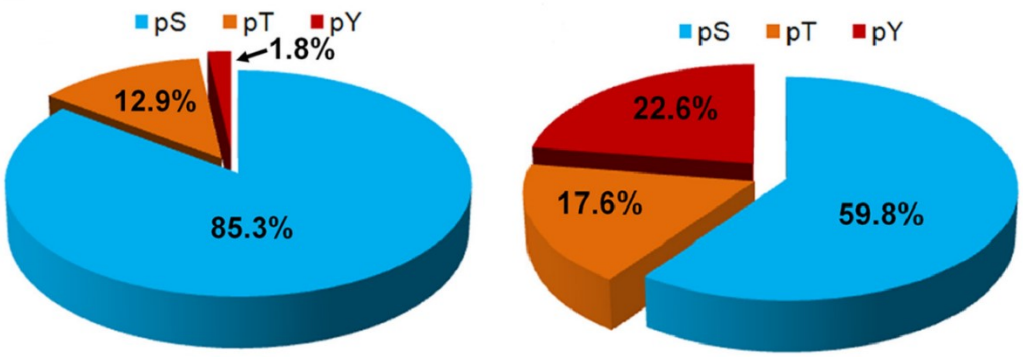

Figure 11: Distribution of phosphorylated serine, threonine and tyrosine in the human liver (left) and in the whole human phosphoproteome (right). The blue part is phosphorylated serine, the orange part is phosphorylated threonine and the red part is phophorylated tyrosine. Adapted from Song et al., „Systematic analysis of protein phosphorylation networks from phosphoproteomic data“ (Song et al., 2012).

In 2008 team of Sugiyama mapped phosphorylation sites in *Arabidopsis thaliana in vivo* using mass spectrometry-based technologies. They identified 20 172 unique phosphorylation sites including 94 tyrosine residues on 1 346 proteins from unfractionated cell lysates. As tyrosine kinases similar to humans were not observed in plants, relative abundances of phosphorylated serine, threonine and tyrosine measured to be 85 %, 10,7 % and 4,3 % were little bit unexpected for *Arabidopsis.* It is suggested that phosphorylated tyrosine residues in plants are likely result of dual-specific serine/threonine/tyrosine protein kinases activity. Further Sugiyama et al. found that most of phosphopeptides containing phosphorylated tyrosine were multiple phosphorylated while the majority of phosphopeptides were singly phosphorylated (Sugiyama et al., 2008).

These results were enriched by Nakagami et al. in 2010, who identified 1 776 novel phosphorylation sites for *Arabidopsis.* They used lactic acid-modified titania, β-hydroxypropanoic acid-modified metal oxide chromatography and Fe(III)-immobilized metal-ion affinity chromatography and estimated the distribution of phosphorylated serine, threonine and tyrosine to be 82,7 %, 13,1% and 4,2 %. Further they identified 5 523 unique phosphorylation sites on 3 393 proteins from non-stimulated suspension-cultured rice cells lysates. For rice they estimated the distribution of phosphorylated serine, threonine and tyrosine to be 84,8 %, 12,3 % and 2,9 % (Nakagami et al., 2010).

5. Effects of phosphorylation on protein structure

Physico-chemical properties, stability, structure, kinetics, and dynamics of protein might be changed by phosphorylation or dephosphorylation on its residue. Adding or removing of phosphate group may modify the intrinsic propensity toward certain local backbone conformations or modify the global interactions due to double-charged nature of the phosphate group or its size. As a result phosphorylation can promote conformational changes at local as well as global levels that can lead to complex binding energy modifications and order-to-disorder or disorder-to-order transitions. Phosphorylation can, further, serve as a recognition site for an enzyme which catalyzes the conformational switch (Nishi et al., 2014). Or eventually in some proteins the removal of phosphate has presumably no effect on their activity and the function of the phosphate is not yet known. This type of phosphorylation is termed as “silent phosphorylation” (Johnson, 1993).

As other posttranslational modifications, phosphorylation can influence physico-chemical properties of proteins in two different ways. Firstly, phosphorylation can occur at the functional site or in its immediate neighborhood. The added phosphate can act ortosterically and can be recognized by recognition domains or block active sites through direct interference with substrate binding. Secondly, upon phosphorylation the atoms around the linked group have to adapt themselves to optimize their interactions. It leads to forming new interactions or breaking the previous. These changes in positions of amino acid residues may propagate across the protein structure and induce conformational changes of protein. In this way, phosphorylation allosterically influences the activity and dynamics of protein (Nussinov et al., 2012).

5.1. Ortosteric effects of phosphorylation

Orthosteric effects can influence a compound binding (binding of small molecules, ligands or peptide substrate) by causing steric hindrance effects, blocking or creating new binding sites for potential compounds.

In 2016 Korkuć & Walther used dataset comprised of 5 250 binding sites for compounds such as ligands, small molecules and peptide substrate and 37 065 phosphorylation sites (PhosphositePlus, dbPTM and Phospho.ELM) located on 4 047 proteins with known three-dimensional structure (from the Protein Data Bank). They analyzed the spatial relationships of

compound binding site on protein surface and phosphorylation site. They found that whereas negatively charged compounds were repelled by phosphate's two negative charges, positively charged compounds were attracted (Figure 12). In addition, neutral compounds were also influenced because of their strong dipole moments that make them susceptible to electrostatic effects. Neutral compounds were attracted to phosphorylated binding site, but less than positively charged compounds. Further they noted that phosphorylation sites were preferentially in clusters in neighborhood of negatively charged compounds. That's why Korkuć and Walther concluded that phosphorylation effects most the binding of negatively charged compounds to proteins (Korkuć and Walther, 2016).

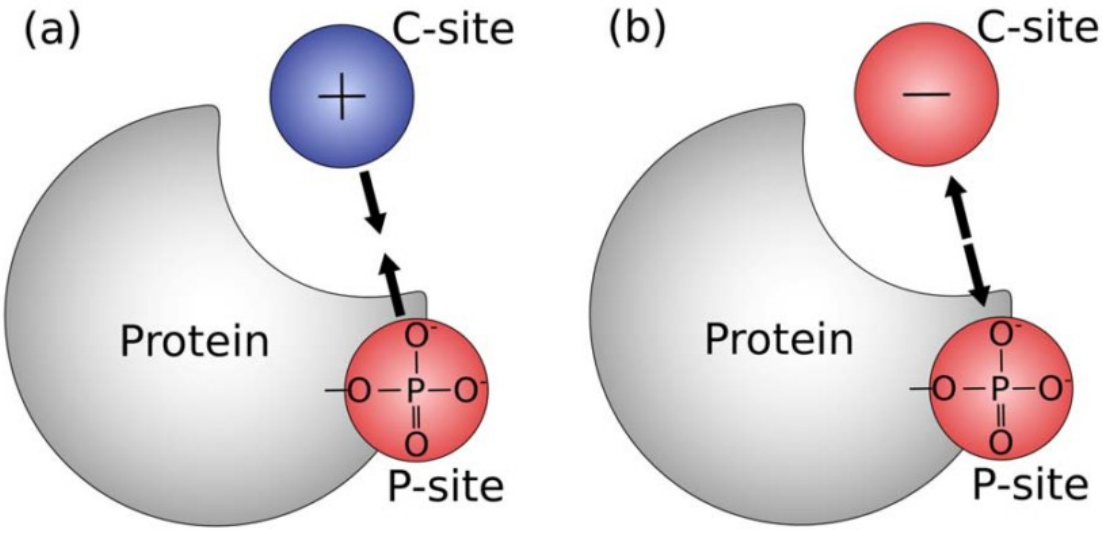

Figure 12: Illustration of influence of phosphate group on compound binding. Because of two negative charges of phosphate group, positively charged compounds are attracted to catalytic site and contrary negatively charged compounds are repelled. Adopted from Korkuć ae at., „Spatial proximity statistics suggest a regulatory role of protein phosphorylation on compound binding" (Korkuć and Walther, 2016).

Phosphorylation may affect domain-motif interaction as an on/off switch. It turns on interaction for domains that are known to interact with motifs only when they are phosphorylated, and it turns off interaction for domains that bind non-phosphorylated motifs (Pawson and Nash, 2003).

In 2012 Akiva et al. studied coupling between motifs and phosphorylation events by analyzing six databases of human experimentally-verified protein phosphorylation sites (Phospho.ELM, PhosphoSite, Uniprot, Human Proteinpedia, phosphorylation site database and Human Protein Reference Database (HPRD). They found that phosphorylation may serve as double-switch for binding domains from different families (namely SH2 and SH3, WW and PDZ

domain), where a phosphorylation event allows one interaction while concurrently preventing another interaction. For example they observed that 150 out of the highly reliable 330 SH2-binding motifs have nearby phosphorylated residue, which is in 60 % of the cases tyrosine. They suggested, that tyrosine can attract tyrosine-kinase to the region and induce the phosphorylation of this site, which is needed for binding of SH2 domain. On the contrary, for SH3 and PDZ-binding motifs phosphorylation may function in opposite way - prevent domain binding (Akiva et al., 2012)(Tatárová et al., 2012).

5.2. Allosteric effects of phosphorylation

Phosphorylated residues in protein act as new chemical entities and therefore diversify the chemical nature of protein surfaces. Addition of a phosphate group that carries two negative charges to a protein can cause conformational change in the protein by attracting cluster of positively charged amino acid side chains and this can alter protein's substrate binding and catalytic activity trough an allosteric effect (Hunter, 2012). For example, phosphorylation at Ser14 within glycogen phosphorylase leads to activation at the catalytic site over 35 Å away and thus phosphorylation causes the activation of the enzyme (Johnson, 1993).

The classic view of allosteric coupling is that two sites are coupled through a network of interactions that extend throughout the protein and connect the two sites. In 2007 Hilser & Thompson made mechanistic model and showed on hypothetical proteins that allosteric coupling is maximized when the domains containing one or both sites are intrinsically disordered - without a single dominant conformational macro-state under physiological conditions (Hilser and Thompson, 2007). It is likely that allosteric site needs to couple to the intrinsic dynamics of the protein, which, in turn, underlies communication with relevant functional sites through coherent collective motions in the whole protein (Guarnera and Berezovsky, 2016). Therefore, when phosphorylation allosterically effects the substrate binding site within protein, it would probably occur in intrinsically disordered regions.

5.2.1. Phosphorylation prefers intrinsically disordered regions

Iakoucheva et al. in 2004 created dataset of phosphorylated and non-phosphorylated sites by extracting 25-residues long sequences centered at serine, threonine and tyrosine sites from the eukaryotic proteins in SWISS-PROT. Than they used phosphorylated sites from PhosphoBase database and removed all sites that had more than 30 % sequence identity

inside the combined data sets. Finally they took disorder-ordered regions characterized by X-ray diffraction, nuclear magnetic resonance and circular dichroism from PDB and analyzed the distribution of the disorder-promoting (arginine, lysine, glutamic acid, proline and serine) and order-promoting (cysteine, tryptophan, tyrosine, isoleucine and valine) residues around phosphorylation sites. They found that phosphorylation sites have in their neighborhood disorder-promoting amino acids, while order-promoting amino acids are less frequent than expected (Iakoucheva et al., 2004).

These results were further supported by Jiménez et al. used 324 nonredundant structural models containing 264 modified serine/threonine and 219 tyrosine residues from mtcPTM database (mtcPTM database had collated since 2004 human and mouse phosphoproteins) and observed that specifically phosphorylation sites globally occur in intrinsically disordered protein region. Only a small proportion (10 % in both human and mouse) of phosphorylated sites were found to be located in structured regions, so majority should be positioned in flexible, unstructured segments and linkers between domains (Jiménez et al., 2007).

Tyanova et al. in 2013 observed enrichment of charged amino acids and depletion of proline, serine and threonine in the surrounding of phosphorylation sites situated in ordered regions. However, they studied the variation of phosphorylation on over 5 000 phosphorylation sites on human HeLa S3 cells during the cell cycle (in phases G1, G1S, Early S, Late S, G2 and M) and compared phosphorylation variation of two groups of sites, sites that reside in ordered regions and sites that lie within disordered regions. They found that over 90 % of the phosphorylated residues (4 675 sites from 5 173; especially multiple phosphorylation sites) occur in disordered regions and exhibit greater phosphorylation changes during the cell cycle than the sites located in structured regions (Figure 13)(Tyanova et al., 2013). So Iakoucheva's and Tyanova's teams observed the same trend of phosphorylation to occur in intrinsically disordered regions by different ways.

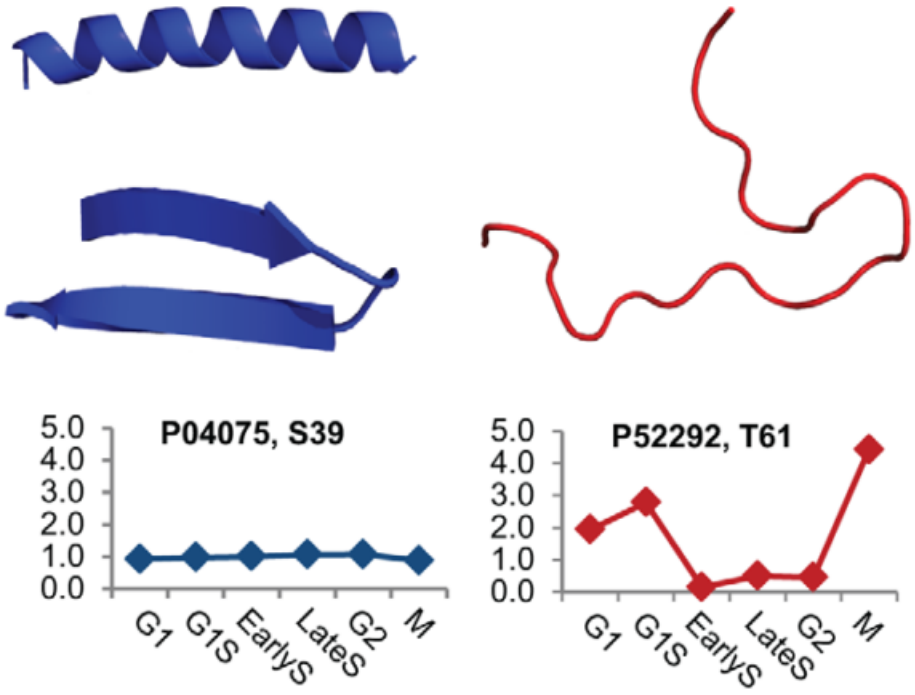

Figure 13: Illustration of non-variable site on a regular secondary structure (α helix and β sheet) and a variable site in disordered regions as measured during the cell cycle. Adapted from Tyanova et al., „Phosphorylation Variation during the Cell Cycle Scales with Structural Propensities of Proteins“ (Tyanova et al., 2013).

Further Nishi et al. in 2013 used dataset of non-redundant human protein complexes from Protein Data Bank and phosphorylation sites from PhosphoSitePlus, Phospho.ELM, and PHOSIDA databases and predicted: 1) intrinsically disordered regions using Disopred and PONDR-FIT programs and 2) whether the phosphorylated residues are a part of an interface between two proteins. They observed strong positive correlation between disorder and phosphorylation on serine, threonine and tyrosine. They also observed weak, but significant correlation between phosphorylation and interface as well as between disorder and interface. 28 % serine, 27 % tyrosine and 20 % threonine residues of all serine, threonine and tyrosine residues in structure are observed on disordered interfaces. Further, while 25 % of serine, threonine and tyrosine residues on disordered interface are phosphorylated, only 8 % of the same amino acids are phosphorylated at ordered interface. In conclusion, these data suggested that phosphorylation is more common in disordered interfaces than in ordered interfaces (Nishi et al., 2013).

Overall, phosphorylation sites seem to prefer intrinsically disordered regions to structured regions. However serine and threonine are different to tyrosine not only in properties of structure but also in preferences where to occur within secondary structure of the protein.

Karabulut & Frishman in 2016 used B-factor analysis on dataset including 609 phosphorylation sites (423 serine, 140 threonine and 46 tyrosine residues) and 10756 non-phosphorylation sites (4162 serine, 3790 threonine and 2804 tyrosine residues) in 332 proteins with known structures. They found that phosphorylated serine preferentially occur in more flexible regions of protein structures. Further phosphorylated serine and threonine residues occur preferentially in loops (Karabulut and Frishman, 2016). According to analysis of mtcPTM database made by Jiménez et al. in 2007 phosphorylated serine and threonine residues are also typically within linkers between domains (structures), whereas phosphorylated tyrosine residues are rather within structured domains, specifically near amino-terminus (Jiménez et al., 2007).

Phosphorylation can even cause order to disorder and disorder to order transitions. These transitions require a balance between the gain in enthalpy (due to electrostatic interactions) and loss in entropy (due to constraining the conformational dynamics of positively charged residues upon phosphorylation). For instance, Bah et al. in 2015 reviewed and showed that intrinsically disordered protein can be folded upon phosphorylation. They observed refolding of the helical motif into β-strand of 4E-BP2 enzyme upon phosphorylation and suggested that refolding may sequester or enhance the accessibilities of protein binding or provide new interaction surfaces (Bah et al., 2015).

5.2.2. Phosphorylation sites are located at exterior surface

Iakoucheva et al. found that phosphorylated serine residues tend to cluster with surrounding serine, lysine, arginine or glutamic acid. They observed both the depletion of cysteine, leucine, isoleucine and aromatic residues nearby serine residues and leucine and isoleucine depletion around threonine and tyrosine sites. In case of tyrosine residues, aspartic and glutamic acid were observed in close proximity to the phosphorylated tyrosine while proline did not occur near the phosphorylated tyrosine. In conclusion, all three sites were both strongly depleted in the neutral and hydrophobic residues and significantly enriched in hydrophilic residues. These results indicated that the position of phosphorylated residues could be near or at surface of protein (Iakoucheva et al., 2004).

The tendency of phosphorylated sites to have less hydrophobic residues in their neighborhood was later affirmed by Frades et al. in 2015. They used a discriminative n-gram analysis on phosphoproteomics datasets (*Toxoplasma gondii, Plasmodium falciparum, Schizo-*

saccharomyces pombe, Saccharomyces cerevisiae, Caenorhabditis elegans, Drosophila melanogaster, Mus musculus, *Homo sapiens, Oryza sativa* and *Arabidopsis thaliana*). They defined the kingdom/phylum-specific phosphorylation motifs. They analyzed the distribution of hydrophobic, negative and positive amino acids as well as the proline content within each kingdom/phyla and found that even though distribution differs slightly between kingdom/phyla, the hydrophilicity of phosphorylated residues and the surrounding sequences is conserved during evolution. It supported the hypothesis, that phosphorylation sites tend to have a local decrease of hydrophobic residues and enrichment in surface exposed residues in order to be highly accessible for the kinases and phosphatases (Frades et al., 2015).

Finally, Karabulut & Frishman in 2016 analyzed structural properties of phosphorylation sites in dataset including 609 phosphorylation sites (423 serine, 140 threonine and 46 tyrosine residues) and 10 756 non-phosphorylation sites (4 162 serine, 3 790 threonine and 2 804 tyrosine residues) in 332 proteins with known structures. They found that phosphorylated serine, threonine and tyrosine along with the residues surrounding them from -10 to +10 are consistently and significantly more solvent exposed (by about 30,73 % for phosphorylated serine, 27,05 % for phosphorylated threonine and 30,02 % for phosphorylated tyrosine) than non-phosphorylated serine, threonine and tyrosine (Karabulut and Frishman, 2016).

Despite the fact, that most side chains of phosphorylated residues tend to be exposed (tyrosine less probably because its large hydrophobic ring), Jiménez et al. found that “nearly 15% of all phosphorylation sites exhibited less than 10% solvent accessibility of their side chains in the unmodified form of the protein”. Phosphorylation of the buried residues may be responsible for structural changes and therefore for regulation of function by affecting functional sites, opening of the structure and increased local flexibility (Jiménez et al., 2007).

5.2.3. Hydrogen bonding and electrostatic effects

Salt bridge occurs between amino acids or other molecules with opposite charges and has some characteristics of hydrogen bonding and electrostatic interactions. Phosphorylated serine, threonine and tyrosine form salt bridges with surrounding water molecules, neighboring backbone groups and/or positively charged side chains of neighboring amino acids. It may impose geometrical constraints, which can modulate the intrinsic propensity toward backbone conformation. Further, the addition of phosphate group, that carries two negative charges, to neutral phosphorylation site may lead to polarization of the electrons on

nearby molecules and to perturbation of the energy landscape of the protein surface, which may result in conformational changes (Mandell et al., 2007). The induction of conformational changes by forming salt bridges between phosphate group and positively charged amino chains is easy to evolve ultra-sensitive switch. In addition, it doesn't require classical cooperativity (Serber and Ferrell, 2007).

Xin & Radivojac studied in 2012 conformational heterogeneity of groups of protein structures corresponding to identical sequences in their nonphosphorylated and phosphorylated forms on dataset obtained from Protein Data Bank. They found that phosphoryl group introduced new hydrogen bonds and salt bridges in the local neighborhood of the modified residue in 64,3 % of cases. Phosphorylation can also change the way how protein interacts with water. A protein that normally doesn't interact with water will become hydrophilic by phosphorylation because of the phosphate group ability to form three or more hydrogen bonds. Moreover, they observed significant local changes in protein conformation (conformational changes bigger than 0,5 Å RMSD occur in 20 % of cases of phosphorylation). In terms of extreme changes, structural changes bigger than 2 Å were measured in 6,6 % of cases. That's why authors suggested that forming of new hydrogen bonds and salt bridges in the local neighborhood of the modified residue can stabilize protein structure (Figure 14) (Xin and Radivojac, 2012).

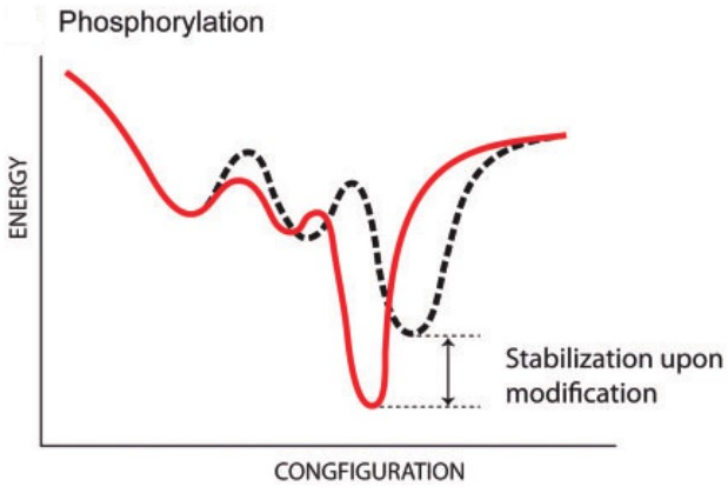

Figure 14: Alteration of the energy landscape induced by phosphorylation. The red curve corresponds to unmodified structures and the dotted black curve shows how energy landscape of post-translationally modified proteins changed is. Adapted from Xin & Radivojac, „Post-translational modifications induce significant yet not extreme changes to protein structure" (Xin and Radivojac, 2012).

Finally, a strength of these bonds varies depending on the identity, proximity, and orientation of the participating side chains and their surrounding environment. In particular, positively charged lysine and arginine can interact with phosphate group and actually phosphate group and arginine achieve more tight binding than do aspartic acid or glutamic acid with arginine, because of the guanidinium moiety planar geometry and ability to form multiple hydrogen bonds to the phosphate oxygens (Mandell et al., 2007).

5.2.4 Salt bridges may alter the backbone conformational preferences of proteins

Posttranslational modifications are generally known to stabilize or destabilize individual secondary structural elements in intrinsically disordered protein regions. Owing to charge-charge interaction of phosphoryl group with sidechains in the local neighborhood, phosphorylated sites appear to be electrostatically stabilized, e.g. in case of neighborhood of basic sidechains (Kitchen et al., 2008).

In 2011 Kim et al. compared phosphorylated and non-phosphorylated peptides in order to find whether or not phosphorylation effects serine and threonine peptide conformations. They carried out pH- and temperature-dependent circular dichroism as well as 1H NMR studies and found that phosphorylation has different effects on the backbone conformational preferences of serine and threonine peptides. They observed that upon phosphorylation serine forms strong hydrogen bonds with surrounding water molecules and intramolecular hydrogen bonds with hydrogen atoms at the acetyl ends, so the phosphate in serine is not available for the intramolecular hydrogen bonds with the hydrogen atoms at the amide ends. Therefore certain part of protein adopt polyproline II conformation. In contrast to this, phosphorylated threonine has additional methyl group at the Cβ-position, so has smaller flexibility of the side chain and thus preferentially makes intramolecular hydrogen bonding with amino groups. It makes the β-strand conformation preferred to polyproline II (Kim et al., 2011).

Further He et al. studied effects of phosphorylation on the intrinsic propensity for serine and threonine by molecular dynamics simulations. Similar to previous findings, He et al. in 2016 found that the phosphorylation of serine enhances the propensity for the polyproline II conformation and phosphorylation of threonine disfavors polyproline II conformation because

of same reasons mentioned previously. On the other hand, He et al. found that phosphorylation tends to decrease the propensity of β-strand formation and promotes the helical conformation for serine as well as threonine peptides (Figure 15) (He et al., 2016).

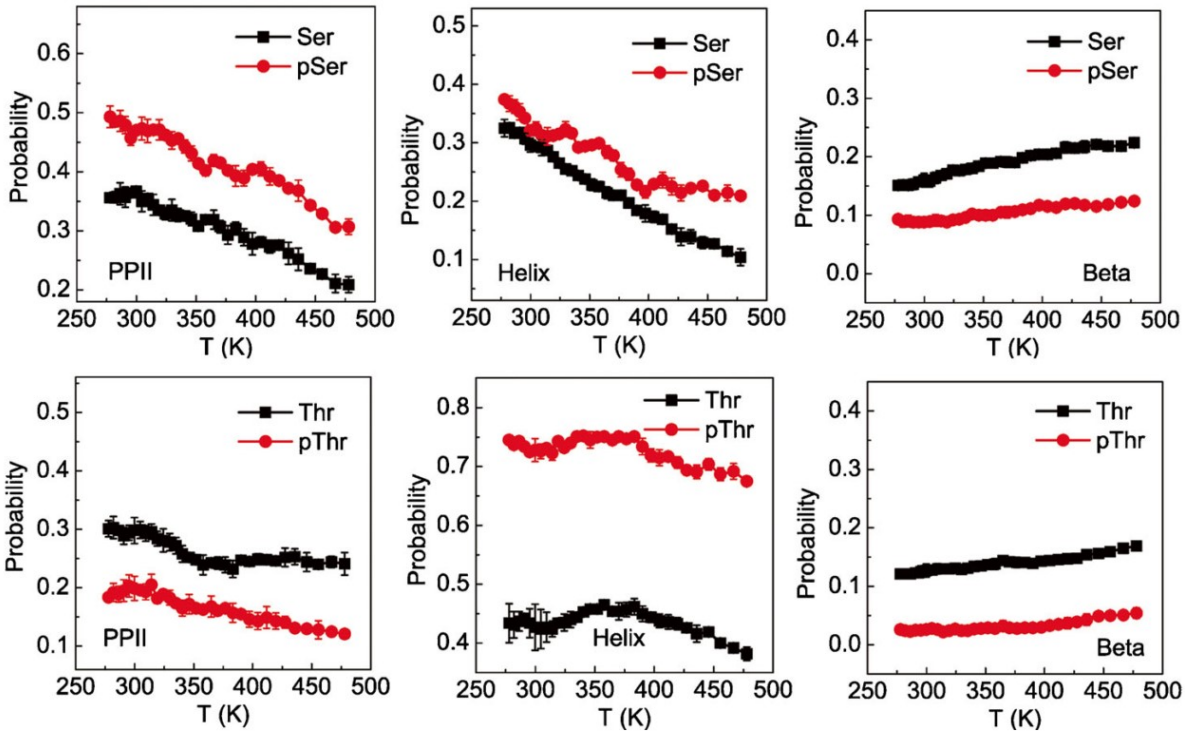

Figure 15: Probability of the polyproline II (left), helix (middle) and β-strand (right) conformations of non-phosphorylated (black curves) and phosphorylated (red curves) serine (upper) and threonine (lower) model peptides. Adopted from He et al., „Effects of phosphorylation on the intrinsic propensity of backbone conformations of serine/threonine“ (He et al., 2016).

In addition, He et al. showed that proportion of amino acids adopting β-strand conformations increases with increasing temperature instead of helix and polyproline II conformations decreases for both serine and threonine peptides. Further He et al. confirmed that serine has weaker interactions with the backbone amino group then threonine after phosphorylation. Overall, phosphorylation of serine and threonine increases the probability of forming hydrogen bond with amino group, stabilizes the helical conformation and moderately destabilize the β-strand conformation (He et al., 2016).

5.2.5 Salt bridges can modify stability of complexes

Nishi et al. in 2011 used a nonredundant set of 933 structures of phosphorylated human hetero- and homooligomeric complexes derived from the PhosphoSitePlus, Phospho.ELM, and PHOSIDA databases and studied locations of phosphorylation sites in complexes and at binding interfaces They found that phosphorylation occur more often than expected by

chance on residues at binding interface of heterooligomers and weak transient homooligomers and suggested that phosphorylated residues form hydrogen bonds at interfaces and therefore contribute to complex stability. However, they also found that the heterooligomers are destabilized in 39 % of cases as well as the homooligomers in 35 % of cases. So phosphorylation leads to destabilization in a third of cases (Nishi et al., 2011).

6. Conclusion

The phosphorylation or dephosphorylation of a protein is a flexible mechanism for reversibly altering its conformation and hence its ability to function. In this thesis recent works that describe and analyze conformational changes in proteins upon phosphorylation in a high-throughput manner are summarized.

This thesis briefly introduces general mechanism of phosphorylation and physico-chemical properties of phosphorylated serine, threonine and tyrosine. Differences such as extra methyl group on the backbone of threonine or the aromatic character of tyrosine explains different behavior of phosphorylated certain residue. In the second chapter of thesis is the conserved structure of protein kinases described and both how the protein kinases and phosphatases regulate many biological processes in cells and are regulated. Moreover, in this chapter is outlined how the components of tyrosine phosphorylation signaling mechanism is evolved shortly before metazoan multicellularity. The third chapter shows the differences in the distribution of phosphorylated residues among species.

The main part of thesis presents how the orthosteric effects of phosphorylation on residues near the binding site affects the substrate binding by the electrostatic impact of negatively charged phosphate group. Whereas positively charged and less the neutral compounds are attracted to the binding site with phosphate nearby, negatively charged compounds are repulsed.

Further, thesis shows the tendency of phosphorylation sites and the residues surrounding them to reside in disordered regions and irregular secondary structures as well as the tendency of phosphorylation sites to be dynamically regulated throughout the cell cycle with the structural features of the sites. It seems to be necessary for allosteric regulation of the protein structure and protein–protein binding, disorder–order and order–disorder coupled transitions upon phosphorylation. Finally, it presents the preference of phosphorylated residues to be located at the protein surface and how its ability to form hydrogen bonds or salt bridges either intra- or inter-molecularly with adjacent positively charged residues or surrounding water molecules may alter the backbone conformational preferences of proteins and complex stability.

The results of analysis on large datasets often disagree (for example the distribution of phosphorylation among serine, threonine or tyrosine), but it is difficult to reconcile the results as they were acquired on datasets of different size and origin by different methods. More data and more analysis will be therefore needed to answer some of the questions presented here. Currently available high-throughput proteomic data, advanced experimental techniques, and computational power give us an opportunity to deepen our knowledge of phosphorylation.

7. References

Akiva, E., Friedlander, G., Itzhaki, Z., and Margalit, H. (2012). A dynamic view of domain-motif interactions. PLoS Comput. Biol. *8*, 1–13.

Apweiler, R., Bateman, A., Martin, M.J., O'Donovan, C., Magrane, M., Alam-Faruque, Y., Alpi, E., Antunes, R., Arganiska, J., Casanova, E.B., et al. (2014). Activities at the Universal Protein Resource (UniProt). Nucleic Acids Res. *42*, 191–198.

Bah, A., Vernon, R.M., Siddiqui, Z., Krzeminski, M., Muhandiram, R., Zhao, C., Sonenberg, N., Kay, L.E., and Forman-Kay, J.D. (2015). Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. Nature *519*, 106–109.

Bodenmiller, B., Campbell, D., Gerrits, B., Lam, H., Picotti, P., Schlapbach, R., and Aebersold, R. (2009). PhosphoPep – a Database of Protein Phosphorylation Sites for Systems Level Research in Model Organisms. Nat. Biotechnol. *26*, 1339–1340.

Cieśla, J., Fraczyk, T., and Rode, W. (2011). Phosphorylation of basic amino acid residues in proteins: Important but easily missed. Acta Biochim. Pol. *58*, 137–148.

Cohen, P. (2000). The regulation of protein function by multisite phosphorylation--a 25 year update. Trends Biochem. Sci. *25*, 596–601.

Dinkel, H., Chica, C., Via, A., Gould, C.M., Jensen, L.J., Gibson, T.J., and Diella, F. (2011). Phospho.ELM: A database of phosphorylation sites-update 2011. Nucleic Acids Res. *39*, 261–267.

Duan, G., and Walther, D. (2015). The Roles of Post-translational Modifications in the Context of Protein Interaction Networks. PLoS Comput. Biol. *11*, 1–23.

Enserink, J.M., and Kolodner, R.D. (2010). An overview of Cdk1-controlled targets and processes. Cell Div. *5*, 11.

Frades, I., Resjö, S., and Andreasson, E. (2015). Comparison of phosphorylation patterns across eukaryotes by discriminative N-gram analysis. BMC Bioinformatics *16*, 1–13.

Gnad, F., Gunawardena, J., and Mann, M. (2011). PHOSIDA 2011: The posttranslational modification database. Nucleic Acids Res. *39*, 253–260.

Goel, R., Harsha, H.C., Pandey, A., and Keshava Prasad, T.S. (2012). Human Protein Reference Database and Human Proteinpedia as Resources for Phosphoproteome Analysis. Mol. Biosyst. *8*, 453–463.

Guarnera, E., and Berezovsky, I.N. (2016). Allosteric sites: Remote control in regulation of protein activity. Curr. Opin. Struct. Biol. *37*, 1–8.

He, E., Yan, G., Zhang, J., Wang, J., and Li, W. (2016). Effects of phosphorylation on the intrinsic propensity of backbone conformations of serine/threonine. J. Biol. Phys. *42*, 247–258.

Hilser, V.J., and Thompson, E.B. (2007). Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. Proc. Natl. Acad. Sci. U. S. A. *104*, 8311–8305.

Hornbeck, P. V., Kornhauser, J.M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V., and Sullivan, M. (2012). PhosphoSitePlus: A comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res. *40*, 261–270.

Huang, K.Y., Su, M.G., Kao, H.J., Hsieh, Y.C., Jhong, J.H., Cheng, K.H., Huang, H. Da, and Lee, T.Y. (2016). dbPTM 2016: 10-year anniversary of a resource for post-translational modification of proteins. Nucleic Acids Res. *44*, 435–446.

Hunter, T. (2007). The Age of Crosstalk: Phosphorylation, Ubiquitination, and Beyond. Mol. Cell *28*, 730–738.

Hunter, T. (2010). Tyrosine phosphorylation : thirty years and counting. Curr. Opin. Plant Biol. *21*, 140–146.

Hunter, T. (2012). Why nature chose phosphate to modify proteins. Philos. Trans. R. Soc. B Biol. Sci. *367*, 2513–2516.

Hunter, T. (2014). The Genesis of Tyrosine Phosphorylation. Cold Spring Harb. Lab. Press *6*, 1–15.

Huse, M., and Kuriyan, J. (2002). The conformational plasticity of protein kinases. Cell *109*, 275–282.

Iakoucheva, L.M., Radivojac, P., Brown, C.J., O'Connor, T.R., Sikes, J.G., Obradovic, Z., and Dunker, A.K. (2004). The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res. *32*, 1037–1049.

Jiménez, J.L., Hegemann, B., Hutchins, J.R. a, Peters, J.-M., and Durbin, R. (2007). A systematic comparative and structural analysis of protein phosphorylation sites based on the mtcPTM database. Genome Biol. *8*, 1–20.

Jin, J., and Pawson, T. (2012). Modular evolution of phosphorylation-based signalling systems. Philos. Trans. R. Soc. B Biol. Sci. *367*, 2540–2555.

Johnson, L.N. (1993). THE EFFECTS OF PHOSPHORYLATION ON THE STRUCTURE AND FUNCTION OF PROTEINS. Annu. Rev. Biophys. Biomol. Struct. *22*, 199–232.

Johnson, L.N., and Lewis, R.J. (2001). Structural basis for control by phosphorylation. Chem. Rev. *101*, 2209–2242.

Karabulut, N.P., and Frishman, D. (2016). Sequence- and Structure-Based Analysis of Tissue-Specific Phosphorylation Sites. PLoS One *11*, 1–19.

Kim, S.Y., Jung, Y., Hwang, G.S., Han, H., and Cho, M. (2011). Phosphorylation alters backbone conformational preferences of serine and threonine peptides. Proteins Struct. Funct. Bioinforma. *79*, 3155–3165.

King, N., Westbrook, M.J., Young, S.L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I., et al. (2008). The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. Nature *451*, 783–788.

Kitchen, J., Saunders, R.E., and Warwicker, J. (2008). Charge environments around phosphorylation sites in proteins. BMC Struct. Biol. *8*, 1–11.

Knighton, D.R., Zheng, J.H., Ten Eyck, L.F., Ashford, V.A., Xuong, N.H., Taylor, S.S., and Sowadski, J.M. (1991). Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. Science (80-.). *253*, 407 LP-414.

Korkuć, P., and Walther, D. (2016). Spatial proximity statistics suggest a regulatory role of protein phosphorylation on compound binding. Proteins Struct. Funct. Bioinforma. *84*, 565–579.

Kreegipuu, A., Blom, N., and Brunak, S. (1999). PhosphoBase, a database of phosphorylation sites: Release 2.0. Nucleic Acids Res. *27*, 237–239.

Lim, S., and Kaldis, P. (2013). Cdks, cyclins and CKIs: roles beyond cell cycle regulation. Development *140*, 3079–3093.

Mandell, D.J., Chorny, I., Groban, E.S., Wong, S.E., Levine, E., Rapp, C.S., and Jacobson, M.P. (2007). Strengths of hydrogen bonds involving phosphorylated amino acid side chains. J. Am. Chem. Soc. *129*, 820–827.

Nakagami, H., Sugiyama, N., Mochida, K., Daudi, A., Yoshida, Y., Toyoda, T., Tomita, M., Ishihama, Y., and Shirasu, K. (2010). Large-Scale Comparative Phosphoproteomics Identifies Conserved Phosphorylation Sites in Plants. Plant Physiol. *153*, 1161–1174.

Nishi, H., Hashimoto, K., and Panchenko, A.R. (2011). Article Phosphorylation in Protein-Protein Binding : Effect on Stability and Function. Struct. Des. *19*, 1807–1815.

Nishi, H., Fong, J.H., Chang, C., Teichmann, S.A., and Panchenko, A.R. (2013). Regulation of protein-protein binding by coupling between phosphorylation and intrinsic disorder: analysis of human protein complexes. Mol. Biosyst. *9*, 1620–1626.

Nishi, H., Shaytan, A., and Panchenko, A.R. (2014). Physicochemical mechanisms of protein regulation by phosphorylation. Front. Genet. *5*, 1–10.

Nishi, H., Demir, E., and Panchenko, A.R. (2015). Crosstalk between signaling pathways provided by single and multiple protein phosphorylation sites. J. Mol. Biol. *427*, 511–520.

Nussinov, R., Tsai, C., Xin, F., and Radivojac, P. (2012). Allosteric post-translational modification codes. Trends Biochem. Sci. *37*, 447–455.

Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006). Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. Cell *127*, 635–648.

Pandya, S., Struck, T.J., Mannakee, B.K., Paniscus, M., and Gutenkunst, R.N. (2015). Testing whether Metazoan Tyrosine Loss Was Driven by Selection against Promiscuous Phosphorylation. Mol. Biol. Evol. *32*, 144–152.

Paul, M.K., and Mukhopadhyay, A.K. (2004). Tyrosine kinase - Role and significance in Cancer. Int. J. Med. Sci. *1*, 101–115.

Pawson, T., and Nash, P. (2003). Assembly of Cell Regulatory Systems Through Protein Interaction Domains. Science (80-.). *300*, 445–453.

Pincus, D., Letunic, I., Bork, P., and Lim, W. a (2008). Evolution of the phospho-tyrosine signaling machinery in premetazoan lineages. Proc. Natl. Acad. Sci. U. S. A. *105*, 9680–9684.

Roskoski, R. (2012). ERK1/2 MAP kinases: Structure, function, and regulation. Pharmacol. Res. *66*, 105–143.

Serber, Z., and Ferrell, J.E. (2007). Minireview Tuning Bulk Electrostatics to Regulate Protein Function. Cell *128*, 441–444.

Song, C., Ye, M., Liu, Z., Cheng, H., Jiang, X., Han, G., Songyang, Z., Tan, Y., Wang, H., Ren, J., et al. (2012). Systematic analysis of protein phosphorylation networks from phosphoproteomic data. Mol. Cell. Proteomics *11*, 1070–1083.

Sugiyama, N., Nakagami, H., Mochida, K., Daudi, A., Tomita, M., Shirasu, K., Ishihama, Y., Altschul, S., Madden, T., Schaffer, A., et al. (2008). Large-scale phosphorylation mapping reveals the extent of tyrosine phosphorylation in Arabidopsis. Mol. Syst. Biol. *4*, 285–286.

Tan, C.S.H., Pasculescu, A., Lim, W.A., Pawson, T., Bader, G.D., and Linding, R. (2009). Positive Selection of Tyrosine Loss in Metazoan Evolution. Science *325*, 1686–1688.

Tatárová, Z., Brábek, J., Rösel, D., and Novotný, M. (2012). SH3 Domain Tyrosine Phosphorylation – Sites , Role and Evolution. PLoS One *7*, 1–8.

Tyanova, S., Cox, J., Olsen, J., Mann, M., and Frishman, D. (2013). Phosphorylation Variation during the Cell Cycle Scales with Structural Propensities of Proteins. PLoS Comput. Biol. *9*, 1–10.

Wortzel, I., and Seger, R. (2011). The ERK Cascade: Distinct Functions within Various Subcellular Organelles. Genes Cancer *2*, 195–209.

Xin, F., and Radivojac, P. (2012). Post-translational modifications induce significant yet not extreme changes to protein structure. Bioinformatics *28*, 2905–2913.